

31/8/1
vef.

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許出願公開番号

特開平6-203000

(43)公開日 平成6年(1994)7月22日

(51)IntCl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/16	4 0 0 U	9190-5L		
13/14	3 1 0 H	8133-5B		
H 0 4 L 12/28		8732-5K	H 0 4 L 11/ 00	3 1 0 Z

審査請求 有 請求項の数 6 (全 26 頁)

(21)出願番号 特願平5-203374

(22)出願日 平成5年(1993)8月17日

(31)優先権主張番号 9 4 6 2 0 4

(32)優先日 1992年9月17日

(33)優先権主張国 米国 (US)

(71)出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州アーモンク (番地なし)

(72)発明者 ハワード トーマス オルノウィッチ
アメリカ合衆国13760、ニューヨーク州エンドウェル、トワイライト ドライヴ 2922

(74)代理人 弁理士 合田 潔 (外6名)

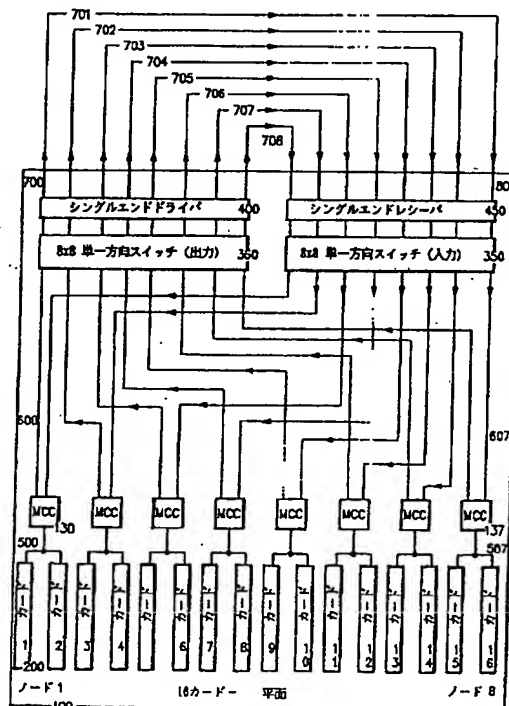
最終頁に続く

(54)【発明の名称】 スイッチ方式マルチノードプレーナ

(57)【要約】

【目的】 複数のプロセッサ(拡張)カードに挿入し、マルチステージスイッチネットワークを介してカードを相互接続するための拡張可能なスイッチ方式平面(プレーナ)を提供する。

【構成】 スイッチ方式平面(100)は、非同期的に複数のノードを接続するためのスイッチネットワークと、各々が個々のバス構造に接続され、カードによって実行される機能に応じて変化する特性のノード要素を構成できる複数のカードスロットグループ(200)と、スイッチネットワークにデータを伝送し、スイッチネットワークからのデータを受信可能な複数のノード要素と、標準的バスアーキテクチャ及びプロトコルをスイッチネットワークのアーキテクチャ及びプロトコルに適用する複数のバスアーキテクチャ変換手段(130, 137)と、ドライバ及びレシーバを含み、平面の外部にある複数のスイッチ経路を相互接続する手段を提供する拡張インタフェース(600, 607)と、から成る。



【特許請求の範囲】

【請求項1】 スイッチ方式マルチノードプレーナであって、

並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続するためのスイッチ相互接続ネットワーク手段と、

各々が個々のバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、

前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、
ドライバ及びレシーバ回路を含み、前記スイッチ方式マルチノードプレーナの外部にある複数のスイッチ経路を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、

から成るスイッチ方式マルチノードプレーナ。

【請求項2】 スイッチ方式マルチノードプレーナであって、

並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続するためのスイッチ相互接続ネットワーク手段と、

各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、

前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスの標準的マルチドロップバスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバスアーキテクチャ変換手段と、

ドライバ及びレシーバ回路を含み、前記スイッチ方式マルチノードプレーナの外部にある複数のスイッチ経路を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、

から成るスイッチ方式マルチノードプレーナ。

【請求項3】 スイッチ方式マルチノードプレーナであって、

並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続して、前記スイッチ方式マルチノードプレーナの内部に配置された他のローカルノード又は前記スイッチ方式マルチノードプレーナの外部に配置された隔離ノードと通信するための分配されたスイッチ相互接続ネットワーク手段の一部と、

各々が個々のバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、

前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、
ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナを相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、

から成るスイッチ方式マルチノードプレーナ。

【請求項4】 スイッチ方式マルチノードプレーナであって、

並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続して、前記スイッチ方式マルチノードプレーナの内部に配置された他のローカルノード又は前記スイッチ方式マルチノードプレーナの外部に配置された隔離ノードと通信するための分配されたスイッチ相互接続ネットワーク手段の一部と、

各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、

前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスのマルチドロップバスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバス変換手段と、

ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナを相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、

から成るスイッチ方式マルチノードプレーナ。

【請求項5】 スイッチ方式マルチノードプレーナであって、

前記スイッチ方式マルチノードプレーナの複数のコピーの中に分配され、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のプレーナに配置された複数のノードを接続するために、前記複数のスイッチ方式マルチノードプレーナの外部で実行された追加のスイッチング要素によって補足される、スイッチ相互接続ネットワーク手段と、

各々が個々の標準的バス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナ及び前記スイッチ方式マルチノードプレーナの外部で実行される追加のスイッチング要素を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成るスイッチ方式マルチノードプレーナ。

【請求項6】 スイッチ方式マルチノードプレーナであって、

前記スイッチ方式マルチノードプレーナの複数のコピーの中に分配され、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のプレーナに配置された複数のノードを接続するために、前記複数のスイッチ方式マルチノードプレーナの外部で実行された追加のスイッチング要素によって補足される、スイッチ相互接続ネットワーク手段と、

各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、

前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、

前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、

各々が単一のマルチドロップバスのマルチドロップバスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバス変換手段と、

ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナ及び前記スイッチ方式マルチノードプレーナの外部で実行される追加のスイッチング要素を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成るスイッチ方式マルチノードプレーナ。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は幾つかの又は多くの計算要素と幾つかの又は多くの入出力要素の少なくとも一方から成るデジタルコンピュータシステム、並びに、スイッチングネットワークを介して並列で互いに高速且つ待ち時間の少ない通信を行う前記個々の要素の能力に関する。

【0002】用語集

カードスロット(Card slots)

プロセッサ又は拡張カードを受信するために平面(プレーナ)に取り付けられるコンセントコネクタ。

拡張カード(Expansion Card)

コプロセッシング(共用処理)又はI/O要素として共に機能する構成要素のボード。カードボードは、ボードに配置された全ての構成要素を相互接続するのに使用される多数の配線平面を有する。カードボードは、カードを平面に接続するのに使用されるコネクタ要素を有する

(通常、カードコネクタは平面コネクタコンセントに挿入される)。カード上のアクティブ構成要素は、通常、コプロセッシングとI/O要素の内の少なくとも一方、遠隔メモリ、並びに、他のプロセッサ又はI/Oカードと通信を行うためカードコネクタを介して平面に接続するマルチドロップバスの形式のI/Oインタフェースを含む。

機能的要素(Functional Elements)

開示される平面に組み込まれるマルチステージネットワークヘプロセッサカード又は拡張カードの形式で接続されるあらゆるプロセッサ又はI/O装置。

I/Oカード(I/O Card)

拡張カードの別名。

MC変換装置(MC Converter)

標準的マイクロチャネルバスプロトコルをスイッチングネットワークプロトコルに変換する機能的変換装置。

マイクロチャネル(MicroChannel)は、殆どのIBM社のパーソナルコンピュータ及び他の機械に使用されている

I/Oマルチドロップバスに対するIBM社の登録商標。明細書において、マイクロチャネル(MicroChannel)はマルチドロップバスとされる。

MCC

MC変換装置の略語。

ノード(Node)

標準的バスによって相互接続される少なくとも1台のプロセッサと拡張カードの少なくとも一方から成り、次にバスがネットワークの単一ポートとしてスイッチングネットワークに接続するシステムの機能的要素。

ノード要素(Nodal element)

ノードに対するもう1つの用語であり、同一の意味をもつ。

平面(Planar)

多数の配線平面を有する相互接続ボード又はマザーボード。平面は、マルチステージスイッチングネットワーク等のアクティブ構成要素及びプロセッサ又は拡張カードを受信することが可能なコネクタスロット等のイナクティブ(inactive)構成要素を含む。平面は前記カード間に相互接続配線を提供する。

ポート(Port)

スイッチネットワークへの単一の双方向入力ポイント又は出力ポイント。

プロセッサカード(Processor Card)

プロセッシング要素として互いに機能する複数の構成要素のボード。カードボードは、ボードに配置された全ての構成要素を相互接続するために使用される複数の配線平面を有する。カードボードは、カードを平面に接続するために使用されるコネクタ要素を有する(通常、カードコネクタは平面コネクタコンセントに挿入される)。カード上のアクティブ構成要素は、通常、計算要素、メモリ(ローカル及びキャッシュ)、及び他のプロセッサ又はI/Oカードと通信を行うためカードコネクタを介して平面と接続するマルチドロップバスの形式のI/Oインタフェースを備える。

受信ノード(Receiving Node)

スイッチングネットワークに標準的バスによって相互接続される少なくとも1台のプロセッサと拡張カードの内の少なくとも一方から成るシステムの機能的要素であり、スイッチングネットワークを介して伝送されるデータを受信する。

送信ノード(Sending Node)

スイッチングネットワークに標準的バスによって相互接続される少なくとも1台のプロセッサと拡張カードの内の少なくとも一方から成るシステムの機能的要素であり、スイッチングネットワークを介してデータを伝送する。

【0003】

【従来の技術】 マルチステージスイッチングネットワークは、最新のデジタル計算システム内で多数の装置を

相互接続するための手段として受け入れられている。特に、並列システムにおいて、N個のシステム要素を相互接続するためにマルチステージ(多段)スイッチングネットワークを使用することは一般的である。ここでNは、幾つかの若しくは数千のプロセッサ又は幾つかの若しくは数千のプロセッサ及び他のシステム要素の組合せであってもよい。

【0004】 最新技術の相互接続方法はマルチドロップバスの回りに集中しており、そのマルチドロップバスは多くの欠点を有し、主なものとして限定された性能(パフォーマンス)及び拡張がある。問題は、よりよいI/Oバス性能及びより多くのI/Oオプションの取り付けに対する必要性が更に増加しているバス方式プロセッサにある。バス方式プロセッサは、より多くのI/Oオプションがバスにタップとして追加されるにつれて性能を失うマルチドロップバス技術の性質と直接矛盾する。概して、マイクロチャネル(MicroChannel)(IBM社の登録商標)等の標準的バスアーキテクチャは性能目的を選択して、その性能レベルにおいて許容可能なI/Oタップの数を制限した。マイクロチャネルの場合は、8個のタップ(又は拡張カード)が200nsサイクル数で生じるバス動作を許容するのに可能なバスタップの最大限の数となる。その結果として、バス方式システムのユーザは、I/O能力がユーザの必要性にかなうには不十分であることを分かり始める。何年もの間、人々はバス方式システムのI/O能力を高めるための手段を探してきた。

【0005】 バス方式システムの性能は、いかなる所定時にも1ユーザだけしかバスへのアクセスを有することができないために制限される。拡張性は、ドライブ能力、ノイズ及び反射を含むマルチドロップバスの電気的特性のために制限される。

【0006】 他方、マイクロチャネル及びバス方式の機械は比較的安価で、プロセッサ性能能力を絶えず高めている。これらの機械及び機械の拡張カードは投資(インベストメント)を要し、所有者は長年の間この投資での使用及び作成を望む。本発明は、新たな平面相互接続手段を開示することによって、高められた性能及び拡張性レベルでバス方式のプロセッサカード及びバス方式の拡張カードを使用する手段を提供する。

【0007】 本発明は、各カード又は複数のカードのセットを高速スイッチングネットワークを介して相互接続させるため、現行のバスインタフェースを適用することによって複数のバス方式カードを相互接続する。これは、個々のバス方式カードに少ない待ち時間のメッセージと通信させ、並列システムとして対話させるのを可能にする。今や多くの通信が並列で同時に行われ、拡張は実質的に無制限である。これは投資の利用を可能にするが、単一バス方式アーキテクチャに置かれた全ての制限を克服する。

【0008】多数の最新技術のスイッチの解決は、スイッチングネットワーク特性及び最新の相互接続システムに必要な少ない待ち時間の概念を提供していない。必要とされる特性は、要素の相互接続を動的且つ迅速に設定及び破壊する能力、1つのチップに安価で且つ容易に設定及び破壊を行う能力、何千もの要素に拡張性を有する能力、あらゆる長さの非校正相互接続配線長をも許容する能力、分配されたクロッキング問題を解決して将来の周波数の増加を可能にする能力、並びに、N個の交換経路を介して並列設定及びデータ伝送を同時に行うのを可能にする能力を含む。

【0009】必要とされる相互接続特性を提供するために本明細書中で使用される分配され、十分に並列なスイッチは、米国特許出願第07/677,543号に開示されるALLNODE スイッチ（非同期、少ない待ち時間、ノード間スイッチ(Asynchronous, Low Latency, inter-NO DE switch)であり、少ない待ち時間及び高い帯域幅で変換されたバスインタフェースの交換を行うために本発明によって適用される。ALLNODE スイッチは高い帯域幅で回路交換能力を提供し、各スイッチ内で分配されたスイッチ経路接続の設定制御及び破壊制御を個々に含み、従って、並列設定し、待ち時間を少なくし、中心点の故障を取り除く。更に、詳細な説明において、ALLNODE スイッチ及び本発明を用いて、バス方式のプロセッサ相互接続問題を効果的に解決するための方法が記載されている。

【0010】しばしばシステムは、性能を高め、フォールトトレランスを提供し、ブロッキングを防ぐためにマルチステージスイッチングネットワークを介する複数の経路を必要とする。本発明によって適用される米国特許出願第07/677,543号で開示されるALLNODE スイッチの発明は複数の経路を提供する。IBM社のフラナセック(P. Franaszek)著、「相互接続ネットワークにおける多重経路階層(Multipath Hierarchies in Interconnection Networks)」では、一方が待ち時間の少ないメッセージ転送を提供し、他方がより長い待ち時間で保証されたメッセージの転送を提供するような、ネットワークへの2つの階層経路が記載されている。メッセージは、最初に待ち時間の少ない経路を介して試行される。ブロッキング又は競合のために伝送が失敗したならば、保証された転送経路を介して再度伝送される。これによって、通常、約90%のメッセージが待ち時間の少ない経路を介して無事に送信され、再伝送のために待ち時間の少ない経路でブロックされたメッセージが転送されることを保証する。

【0011】米国特許第4,952,930号では、第2のバッファされた経路を使用した方法を記載しており、現行の方法に幾分似ている。しかしながら、その方法を実行するのに複数のスイッチの要求によって苦しめられた。米国特許第4,952,930号の教示を採用

するのを妨げるものはないが、より簡単且つより柔軟な方法がマルチステージネットワークを生成するのに必要とされ続けていた。

【0012】マルチステージネットワークは、コンピュータシステム内の複数の装置を相互接続するための受容された手段となっている。マルチステージネットワークは、従来のクロスバー相互接続の置換となる。クロスバーは、依然としてネットワーク相互接続の最も効率的な方法だが、大型システムでは非実用的な傾向がある。N x Mクロスバーは全体の相互接続を許容し、ここで全てのN装置はM装置のセットの異なるメンバと同時に通信を行うことが可能である。いかなる所定のN装置が、遊休(IDLE)であるM装置（いかなる他のN装置とも接続されない）と相互接続するのを妨げるようなものがクロスバーの内部にないために、クロスバーは「非ブロッキング(non-blocking)」である。N装置が使用中(BUSY)であるM装置（以前は他のN装置と接続された）と接続するのを望むならば、以前の接続が破壊されるまで接続できない。しかしながら、これは「競合(contention)」と称されるのであって、「ブロッキング(blocking)」とは称されない。

【0013】N及びMの数が多くなると（通常、32又は64より多い）、複雑さがN x Mの割合で増加し、ピンカウントが(N x M) x Wの割合で増加するために、クロスバーを作成するのはあまり適切ではない。ここでWはポートあたりのピンの数である。従って、大型ネットワークは、通常、より小さなクロスバーの幾つかのステージを互いに連結(カスケード)することによって構成されるマルチステージネットワークから作成されて、拡張されたネットワークを提供する。マルチステージネットワークの不利な点は、マルチステージネットワークが「ブロッキング(blocking)」、即ち、遊休(IDLE)装置に必要な接続を行うネットワークに使用可能な経路がないために、遊休(IDLE)M装置に接続を行うことができないことである。

【0014】参照できる他の特許として米国特許第4,914,571号があり、アドレス指定の方法によって、ネットワークに接続される資源を発見する方法を記載しているが、実際のネットワーク自体のハードウェアを扱っていない。

【0015】米国特許第4,455,605号はバス指向システムについてであり、マルチステージネットワークについてではない。同様に、米国特許第4,396,984号は、I/Oバスチャネルについてであり、マルチステージネットワークについてではない。米国特許第4,570,261号は、バス指向システムでの故障回復についてであり、マルチステージネットワークについてではない。

【0016】米国特許第4,207,609号はI/Oバスチャネルを示して、当業者は本発明の内容との間の

相違を理解する。米国特許第4、207、609号はマルチステージネットワークについてではない。

【0017】米国特許第4、873、517号は、全く異なるタイプのネットワークについてであり、本明細書に記載されるような等距離のマルチステージネットワークについてではない。また米国特許第4、932、021号は、コンピュータボックス内のバス配線経路についてであり、マルチステージネットワークについてではない。米国特許第4、733、391号は、リング相互接続ネットワークを示し、マルチステージネットワークとは異なる。米国特許第4、811、201号は、マルチステージネットワークに適用できない。米国特許第4、754、395号は、リング相互接続ネットワークについてである。

【0018】

【発明が解決しようとする課題】本発明の目的は、複数のバス方式プロセッサカードと複数の拡張カードの少なくとも一方に挿入し、本発明の平面上にあるマルチステージスイッチネットワークを介して前記カードを相互接続するためのモジュール的に拡張可能なスイッチ方式平面装置を提供することである。

【0019】

【課題を解決するための手段と作用】本発明は、複数のバス方式プロセッサカードと複数の拡張カードの少なくとも一方に挿入し、本発明の平面上にあるマルチステージスイッチネットワークを介して前記カードを相互接続するためのモジュール的に拡張可能なスイッチ方式平面装置である。マルチステージネットワークは、バス方式カード（標準的マイクロチャネルバス及びコネクタを提供するマイクロチャネルカード等のバスインタフェースを提供するカード）を相互接続するために使用される。スイッチングネットワークが平面に作成される。

【0020】カード自体はいかなる修正や変更も必要とせず、平面への接続は平面が標準的マルチドロップバス相互接続を含むように作成される。しかしながら、開示される平面は、標準的マルチドロップバス接続を支援するよりもむしろ、複数のスイッチステージを相互接続するための2地点間（ポイントツーポイント）配線しか含まない新たな概念である。全てのプロセッサ及び拡張カードは依然として完全に相互接続され、あらゆる2つの要素の間での通信又はメッセージを送信することが可能である。通信媒体はスイッチ方式であり、完全に並列で、 n 個の伝送を同時に支援する。ここで、 n はスイッチングネットワークによって相互接続されるノードの数である。

【0021】好ましい交換手段は、米国特許出願第07/677、543号の「All-nodeスイッチ、非クロック、バッファ不使用非同期交換装置(All-Node Switch, An Unclocked, Unbuffered Asynchronous Switching Apparatus)」に開示される基本的なデジタル非同期及び

バッファ不使用交換概念である。Allnode 交換装置は、実施を簡素化する特徴を有するスイッチングネットワーク通信構造を提供し、いかなる種類のデータバッファリング又はデータ変換も必要としない。装置は、接続を即時に設定又は破壊（切断）し、従って、非常に迅速な応答時間で動的に変更可能である。装置は、要求を分析して並列方法で接続を設定する能力を有し、ここで n 個の接続は同時に設定又は破壊される（ここで、 n はスイッチネットワークを介して通信を行うシステムの要素の数）。従って、同時に作成又は破壊することが可能な接続の数はシステムの大きさに直接比例する。この能力によって、本発明の装置が複数の短いメッセージを非常に効率的に処理することが可能になる。更に、新たな装置は同期要求又は配線長制限がない。また、本装置は本発明の装置を実行するために使用される技術が向上するにつれて、新しい技術の速度の向上をたどり、性能を高める能力も有する。更に、Allnode 交換装置は他の同一の装置と連結（カスケード）されて、あらゆる数のシステム要素又はノードの中で相互接続ネットワークを形成することが可能である。前記ネットワークは完全な並列相互接続に対してこれらの特徴を有する。

【0022】スイッチ方式マルチノードプレーナは、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続するためのスイッチ相互接続ネットワーク手段と、各々が個々のバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、ドライバ及びレシーバ回路を含み、前記スイッチ方式マルチノードプレーナの外部にある複数のスイッチ経路を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

【0023】スイッチ方式マルチノードプレーナは、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続するためのスイッチ相互接続ネットワーク手段と、各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送す

ることができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、各々が単一のマルチドロップバスの標準的マルチドロップバスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバスアーキテクチャ変換手段と、ドライバ及びレシーバ回路を含み、前記スイッチ方式マルチノードプレーナの外部にある複数のスイッチ経路を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

【0024】スイッチ方式マルチノードプレーナは、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続して、前記スイッチ方式マルチノードプレーナの内部に配置された他のローカルノード又は前記スイッチ方式マルチノードプレーナの外部に配置された隔離ノードと通信するための分配されたスイッチ相互接続ネットワーク手段の一部と、各々が個々のバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナを相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

【0025】スイッチ方式マルチノードプレーナは、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のノードを接続して、前記スイッチ方式マルチノードプレーナの内部に配置された他のローカルノード又は前記スイッチ方式マルチノードプレーナの外部に配置された隔離ノードと通信するための分配されたスイッチ相互接続ネットワーク手段の一部と、各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能

なノード要素と、各々が単一のマルチドロップバスのマルチドロップバスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバス変換手段と、ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナを相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

【0026】スイッチ方式マルチノードプレーナは、前記スイッチ方式マルチノードプレーナの複数のコピーの中に分配され、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のプレーナに配置された複数のノードを接続するために、前記複数のスイッチ方式マルチノードプレーナの外部で実行された追加のスイッチング要素によって補足される、スイッチ相互接続ネットワーク手段と、各々が個々の標準的バス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、各々が単一のマルチドロップバスの標準的バスアーキテクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のバスアーキテクチャ変換手段と、ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナ及び前記スイッチ方式マルチノードプレーナの外部で実行される追加のスイッチング要素を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

【0027】スイッチ方式マルチノードプレーナは、前記スイッチ方式マルチノードプレーナの複数のコピーの中に分配され、並列に入力ポートから出力ポートへの接続を介して非同期的に複数のプレーナに配置された複数のノードを接続するために、前記複数のスイッチ方式マルチノードプレーナの外部で実行された追加のスイッチング要素によって補足される、スイッチ相互接続ネットワーク手段と、各々が個々のマルチドロップバス構造に接続され、カードスロットに差し込まれるさまざまなプロセッサと拡張カードの少なくとも一方によって実行される機能に応じて変化する特性のノード要素を構成することが可能な複数のカードスロットグループと、前記スイッチ相互接続ネットワーク手段にデータを伝送することができる複数の差込み可能なノード要素と、前記スイッチ相互接続ネットワーク手段からのデータを受信することができる複数の差込み可能なノード要素と、各々が単一のマルチドロップバスのマルチドロップバスアーキ

テクチャ及びプロトコルを、前記スイッチ相互接続ネットワーク手段のアーキテクチャ及びプロトコルに適用する能力を備える、複数のマルチドロップバス変換手段と、ドライバ及びレシーバ回路を含み、複数の前記スイッチ方式マルチノードプレーナ及び前記スイッチ方式マルチノードプレーナの外部で実行される追加のスイッチング要素を相互接続する手段を提供する前記スイッチ相互接続ネットワーク手段への拡張インタフェースと、から成る。

[0028]

【実施例】図面をより詳細に参照すると、図1はスイッチ方式マイクロチャネル平面(Switch-based MicroChannel Planar)を介してn個のノードを相互接続するための好ましい方法を示している。システム全体は、個々のマイクロチャネル変換装置(MCC330、332、334、336、338、340、342、344、346、348等)に接続される個々のノード(302、304、306、308、310、312、314、316、318、320等)から成り、マイクロチャネル変換装置はスイッチネットワーク30に接続される。一般的なノードは、2つのマイクロチャネルカード(MC CARDS)から成るブロックによって示される。これらのカードはいかなる組合せのプロセッサカード又は拡張カードであってもよい。ノードあたりのマイクロチャネルカードの数は、各個々のカードと共に変化し、1乃至8つのカードで構成可能である。しかしながら、図1に示される好ましい実施例では各ノードに2つのMCカードを割り当てている。これらMCノードは、ノードが1つのMCカードだけで構成される場合に関連するMCC330への直接2地点間配線マイクロチャネルバスとして接続されるか、若しくは、MCC330へのマルチドロップマイクロチャネルバスとして接続されて、ノードに高められたカードカウント及び高められたノード柔軟性を与える。単一のMCC330機能へのノードあたり2つのカードのマイクロチャネルバス接続は、図1においてMC1とラベル付けされた固有のマイクロチャネルバスとして示され、MC2、MC3等の他の同様のMCバスと全く異なり、隔離されたマイクロチャネルバスであることを示している。次に、各MCC機能ブロックは2地点間配線によってスイッチネットワーク30だけに接続する。

【0029】スイッチネットワーク30は、幾つかの又は多数のノードが相互接続するのを許容することによって小さな又は大きな拡張を許容するあらゆるタイプのマルチステージ、トーラス、ハイパーキューブネットワークであってもよい。しかしながら、交換手段の好ましい実施例は、米国特許出願第07/677、543号に開示される基本的なデジタル非同期及びバッファ不使用の交換概念であり、単一方向のスイッチの実行である。好ましい実施例は単一方向のインタフェースのみを使用

し、従って、図1は、1本がノード1からのデータを搬送し、1本がスイッチネットワークからMCC1ヘデータを搬送し、1本がMCC1からスイッチネットワークヘデータを搬送する、2つの単一方向のラインのセットから成るスイッチインタフェース70を示している。

【0030】図2乃至図8は米国特許出願第07/677、543号と共通の説明であり、本出願はその内容の全てに関して参照によって本発明に組み込まれることが理解される。図2乃至図8は、交換概念の原理及び速度を説明するALLNODE スwitchの4x4クロスバーの実施例を参照する。

【0031】図2を参照すると、好ましい実施例は4x4のスイッチ装置12であり、ここで本発明の機能は、デジタル、アナログ、光学入力はいかなる4個のセットも相互に排他的な方式で未使用の4個の出力ポートの内のいかなる1個のポートに接続するような手段を提供することである。4x4スイッチ装置12は、いかなる所定時にも4個までの同時接続を支援することができる。例えば、入力1は出力3に接続され、入力2は出力4に接続され、入力3は出力2に接続され、入力4は出力1に接続されることができる。

【0032】本発明のスイッチ装置12は単一方向であり、これはデータが前記スイッチ装置12を介して入力から出力への一方向のみにフローすることを意味する。スイッチ装置12のインタフェースは図2に詳細に画定されている。スイッチ装置12への各入力ポートにおけるラインの集合31、32、33及び34は数が同じで、各出力ポートにおけるラインの集合41、42、43及び44に機能する。各入力ポート及び出力ポートへのインタフェースラインのセットは、4本のデジタルデータライン及び3本のデジタル制御ライン(有効(VALID)、拒否(REJECT)及び受容(ACCEPT))の7個の独自の信号を含んでいる。各ポートにおける信号は、関連するポート(X)の方向及び番号を示す接頭部INX-又はOUTX-によって区別される。4本のデジタルデータライン及び1本の有効(VALID)ラインはスイッチ装置12を介して入力から出力へと進む方向の信号フローを有する。一方、デジタル拒否(REJECT)制御ライン及びデジタル受容(ACCEPT)制御ラインは逆方向の信号フローを有する。

【0033】各単一方向のスイッチインタフェースの集合は、図2に示されるように、ネットワーク30を介するデータを伝送及び制御するために7個の信号しか必要せず、デジタルデータ及び制御転送幅は一度に1/2バイト(4ビット)である。必要とされる信号は以下の通りである。

データ(DATA): スwitch接続を命令し、デジタルデータメッセージ又はデジタル制御ヘッダーを伝送するために使用される4個の並列信号。

有効(VALID): 作動時には、デジタルメッセージ、制

御ヘッダー又はアナログ波形が伝送中であることを示す。作動しない時には、リセット(RESET) コマンドを示して、全てのスイッチを遊休(IDLE)状態にリセットさせる。

拒否(REJECT): 信号フローがデータ(DATA) 信号及び有効(VALID) 信号と逆方向にある。作動時には、拒否(REJECT)、即ち、エラー状態が検出されたことを示す。

受容(ACCEPT): 信号フローが拒否(REJECT) 信号と同一方向にある。ロー状態にあると、メッセージが受信され、正確さが検査されている最中であることを示す。作動時には、メッセージが正確に受信されたことを示す。

【0034】図3によって示されるように、ALLNODE スイッチング装置は複数の入力ポート及び出力ポートを有するノードに対して設けられ、各入力ポートへの接続制御回路、1個の入力をいかなる2個の出力にも接続するための各出力ポートへのマルチプレクサ制御回路から成る。ここで、親出願のように、1及び2は2以上のいかなる固有値であってもよいと仮定できる。

【0035】図3は、4x4クロスバーALLNODE スイッチ装置を示し、ここでALLNODE スイッチ装置10は単一方向であり、これは、データが前記スイッチ装置10を介して入力から出力への1方向だけにフローすることを意味する。前記スイッチ装置10は単一方向だが、図3に示されるように4x4ALLNODE スイッチ装置10を接続することによって4個のノード(20、22、24及び26)の中の双方向通信を支援する。各ノード20、22、24及び26は、1つはスイッチ装置10へ行き、もう1つはスイッチ装置10から来るような2セットの単一方向の相互接続配線を有する。スイッチ装置10の内部の点線は、前記スイッチ装置の機能が入力ポート(INPUT PORT) 1等の入力ポートを4つの可能な出力ポートの内の1つに接続することを示している。スイッチ装置10は、各入力ポートに同一の機能を正確に提供して、いかなる未使用の出力ポートにも接続させることができる。

【0036】図4を参照すると、8個のスイッチ装置10ブロックを連結することによってシステムにおけるノードの数を増加する方法が示されている。8個の連結されたスイッチは10A乃至10Hと示され、入力ポート及び出力ポートの配線に関してのみ変化するスイッチ装置10の同一のコピーであることを示している。あらゆる16個のノードが、スイッチ装置10のブロックの内のちょうど2ブロックを通過する接続によって他のいかなるノードにも通信を行うのが可能なことが注目される。例えば、ノード5は横断スイッチ10B及び横断スイッチ10Hによってノード15にメッセージを送信することができる。全ての接続が2つのスイッチ装置10のブロックを介して作成されるため、8個のスイッチ装置10のブロックから成るネットワークは2ステージスイッチネットワークと称される。他のマルチステージネ

ットワークは、3つのステージ、4つのステージ等を同様に用いてスイッチ装置10のブロックから構成される。従って、スイッチネットワーク30の大きさを変えて、拡張された数のスイッチネットワークポートに相互接続するためにより多くのPC及びワークステーションの少なくとも一方を追加するだけで、このネットワークはノードとしてPC及びワークステーションの少なくとも一方から成る並列システムの大きさをモジュール的に増加するため使用される。

【0037】図5を参照すると、ALLNODE スイッチ装置10を介する単純なデータフローの機能的な図が示されている。スイッチ内の各入力ポート及び各出力ポートにおける有効(VALID) ライン及び4本のデータラインは、簡素化のために図5の単一ラインによって表される。例えば、入力ポート(IN PORT) 1においてスイッチ装置10に入る有効(VALID) ライン及び4本のデータラインは、スイッチ装置10の5個の内部の機能的ブロック50A、60A、60B、60C及び60Dに進む。ブロック50Aは、4つの可能な出力ポートの内のどの1つが入力ポート1に接続されるべきかを決定する。各入力ポートからの有効(VALID) ライン及び4本のデータラインは、各出力マルチプレクサブロック(60A、60B、60C及び60D)に進む。これによって、全ての入力ポートを全ての出力ポートに接続することが可能になる。4個の出力マルチプレクサブロック(60A、60B、60C及び60D)の各々が、制御ブロック(50A、50B、50C及び50D)の各々から、4つの可能な入力ポートラインのセットの内のどれが各出力ポートを介してゲートされるかに関して独自に命令される。例えば、制御ブロック50Aはマルチプレクサ60Cに命令して、入力ポート1を出力ポート3に接続させる。制御ブロック50Bはマルチプレクサ60Aに命令して、入力ポート2を出力ポート1に接続させる。また、制御ブロック50Cはマルチプレクサ60B及び60Dに命令して、入力ポート3をマルチキャスト方法で出力ポート2及び出力ポート4に接続させる。3つの全ての接続は、同時に又は異なる時間に設定できる。マルチプレクサ60A乃至60Dが接続を行って、入力ポートから出力ポートへの単一方向の信号フローでスイッチ装置10を介して有効(VALID) 信号及びデータ信号を移動すると同時に、マルチプレクサ61D及びANDゲート63Dは、出力ポートから入力ポートへと逆方向の信号フローでそれぞれ拒否(REJECT) 信号及び受容(ACCEPT) 信号の信号接続を行う(一般的な実施例はブロック61D及び63Dによって示され、同様のブロックが各入力ポートに関連する)。これらの拒否(REJECT) 信号及び受容(ACCEPT) 信号は、スイッチ装置10に、連結されたネットワークにおける後続のスイッチ装置10のステージによって、或いは、有効(VALID) 信号及び4つのデータ信号を受信及び解釈する装置によって行われる動作の正

のフィードバック表示を行う。有効(VALID)信号の制御下で4つのデータ信号によってスイッチ装置10を介して伝送される制御ヘッダー又はデジタルメッセージは、命令された接続を行うことができないならば全てのネットワークステージによって、又は、この時点でメッセージを受信することができない若しくは伝送中にエラーを検出するならば受信装置によって拒否(REJECT)される。受信装置は、受容(ACCEPT)信号をパルス発生させることによって、(エラーが検出されずに)コマンド又はメッセージの正確な到着を確認することもできる。拒否(REJECT)信号及び受容(ACCEPT)信号がデータフローと逆方向に進むため、信号は試みられた伝送が正確に受信されたか又は拒否されたかの正の表示を送信側に報告する手段を提供する。

【0038】図6を参照すると、ブロック56、52及び54は、スイッチ装置12の部分的な図であるスイッチ装置14に伝送され、スイッチ装置14を介することが可能なメッセージの形式でマルチライン(並列)/直列デジタルデータを生成するための一般的な方法を示している。ブロック56、52及び54によって提供される同様の並列/直列データ生成ロジックは、スイッチ装置12への他の入力ポートの各々で使用する事ができる。入力データラインの各セットは、同一のクロック信号(図6において100MHZ)によって制御されるとして4本の同期化されたデータのライン31をシフトすることによって、並列/直列データを生成する4つのシフトレジスタ54によって同一のクロックに同期化される所定の入力ポートに4本のデータラインを提供する。しかしながら、スイッチ装置14への4つの異なる入力ポートソース(31、32、33及び34)は、異なる非同

同期化された100MHZのクロック信号を基にして互いに非同同期にすることができる。

【0039】スイッチ装置14を介して並列/直列メッセージを送信するための処理は、伝送されるべきデータメッセージを累積するFIFO(先入れ先出し)56を含む。伝送されるべき次の全体のメッセージはバッファ52に移動される。バッファ52に記憶されるメッセージは伝送の準備のためにシフトレジスタ54に移動され、データは4つのシフトレジスタ54を介して分散される。データビット0をシフトレジスタ1の第1ビットに配置し、データビット1をシフトレジスタ2の第1ビットに配置し、データビット2をシフトレジスタ3の第1ビットに配置し、データビット3をシフトレジスタ4の第1ビットに配置し、データビット4をシフトレジスタ1の第2ビットに配置する等によって分散される。次に、並列/直列データがメッセージ全体が伝送されるまで連続してフローするように、シフトレジスタ54は4本の同期化されたデータラインを介してスイッチ装置14に直列データを送信し始める。スイッチ装置14は、(インタフェース31を介する直列レジスタ54からス

イッチ装置14への直列データの最初の2クロックサイクルにおける)伝送される最初の8ビットを使用し、スイッチ装置14を介して接続経路を選択及び設定する。図6の例では、入力ポート1(31)と出力ポート2(42)の間の一時的な接続を行うスイッチ装置を点線で示して、インタフェース31の8本の個々のラインの各々はインタフェース42の対応するラインの各々に独自に且つ直接接続される。

【0040】図7を参照すると、スイッチング装置14の入力ポート及び出力ポート双方の一般的な直列波形が示されている。スイッチ装置は、シフトレジスタ54によって送信されるときに直列伝送の最初の8個のビットを除去し、インタフェース31からインタフェース42への接続等の接続を作成及び保持するために使用する。本例における残りの直列メッセージは、インタフェース31が受信する全く同一のメッセージから最初の8ビットを差し引いたものを見るようにインタフェース31からインタフェース42に直接転送され、直列データがスイッチ装置14を横断するとき直列データによって遭遇する回路遅延によって遅延される。スイッチ装置14は、いかなる方法でもインタフェース31を介して入力する直列データをバッファ又は再クロックしない。スイッチ装置14は、最初の8個のビットを除去する以外はいかなる方法でも波形を変えずに、インタフェース31を介して受信する入力波形をできるだけ速く出力インタフェース42に反転するだけである。

【0041】インタフェース(インタフェース31等)を介してスイッチ装置14の入力ポートにいかなる転送も行われていないことを示す規則は、4個のデータライン及び有効(VALID)制御ラインが論理0に保持されていることを示す連続した遊休(IDLE)コマンドを出すことである。いかなる入力ラインの論理1の検出によっても、遊休(IDLE)状態から出て、スイッチ装置に選択及び転送が開始していることが示される。同様に、スイッチ装置からの出力ラインは、アクティブ転送が行われていないとき、遊休(IDLE)状態(全てが0に)に保持される。

【0042】概して、全てのスイッチ装置が経路選択方法を必要とすることによって、スイッチ装置はどの接続(入力ポートから出力ポートへ)を行うかが命令される。スイッチ装置10において、経路選択コマンドはデータが伝送されるのと同じインタフェース、即ち、各入力ポートに関連する4本のデータライン、を介してスイッチ装置に伝送される。選択情報は、命令された相互接続が行われてデータが命令された宛先にフローすることができるようデータ転送の前に転送されなければならない。選択情報は、特定の入力番号を介してスイッチに到着し、スイッチがどの入力番号がデータを受信しているかを既に知っているために、入力ポート番号(1乃至4)を識別する必要はない(NOT)。従って、選択情報はスイッチ装置10の4つの出力ポートの内のどの1つを

どれに接続するか番号(1乃至4)を特定するだけ(0 NLY)である。本明細書で薦める経路選択の方法は、ゼロ復帰にコード化しているN個の内の1個である(DEAD FIELD と称される)。

【0043】図8を参照すると、正確な直列ビットパターン及び制御信号作動化の一般的な例が、スイッチ装置10に制御及びデジタルデータ情報を送信するために示される。本例は、図4に示される連結された2ステージスイッチネットワークを参照し、ノード1からスイッチ装置10A及び10Fを介するノード7へのネットワークを介してデータを送信することを含む。この接続を行うため、入力ポート1は第1ステージスイッチ装置10Aの出力ポート2に接続されなければならない、入力ポート1は第2ステージスイッチ装置10Fの出力ポート3に接続されなければならない。図4にて、スイッチ装置10A及び10Fに所望の接続をさせるために入力ポート1に送信される信号シーケンスが示されている。1及び0の信号シーケンスにおいて、時間は左から右へと進み、クロックタイム-2で見られる値はスイッチ10Aに第1に到着し、クロックタイム-1の値は第2に到着する等となる。ラインが遊休(IDLE)と示すために、タイム-2及び-1の間でINI-DATAライン及びINI-VALIDラインの値は全てゼロであり、スイッチ10Aに何も起こらない。クロックタイム0において、INI-VALIDラインの値が論理1になる。これは入力ポート1にデータを受信させることによってスイッチ10Aを準備させるが、この時点ではスイッチ10Aにいかなる接続や動作も起こらない。INI-VALID 制御ラインは、基本的に対応するスイッチ入力ポートを使用可能にする。INI-VALIDラインが論理0であるとき、スイッチ10Aがあらゆる接続もできないか、或いは、リセット(RESET)に保持される入力ポート1からいかなるデータも受信できない。最後に、クロックタイム1において、スイッチ10Aはどの出力ポートに接続するかに関するコマンドを受信する。コマンドはクロックタイム1の間に全て受信される。

【0044】クロックタイム1で送信されるコマンドビットパターンは、出力ポートへの接続を行うためにスイッチ10Aによって使用される。このプロセスは経路選択動作と称され、スイッチ10Aの内部で全て行われる。本発明のALLNODE スイッチによって実行される経路選択方法は、4本のINI-DATAラインの各々に選択されるべきスイッチ10Aの独自の出力を画定させる。例えば、タイム1で論理1に進むINI-DATA1 信号はスイッチ10Aに出力ポート1に接続するよう命じ、INI-DATA2は出力ポート2に接続するよう命じる等となる。本例では、INI-DATA2 がクロックタイム1の間に論理1に進むため、従って、スイッチ10Aは出力ポート2に接続するよう命じられる。換言すれば、接続アルゴリズムとは、入力ポートが使用可能にされた後に論理1に進む第1データ入力ラインがその入力ポートが作成すべき接続

を画定することである。これは、通常の場合、クロックタイム1の1本のデータラインのみが論理1にされ、他の3本のデータラインが論理0でなければならないという点で相互に排他的なプロセスである。1つのビットの選択情報が論理1に保証されるならば、転送が開始していることを示すいかなる追加のビットも必要とせずにスイッチ10A伝送の開始を認識することができることに注意したい。スイッチ10Aは、データラインから4つのビットを除去し、図5の制御ブロック50Aの選択レジスタに4つのビットを記憶することによって命令された選択を行う。クロックタイム1の間に伝送されるビットはスイッチ10Aからスイッチ10Fへと渡されず、代わりにスイッチ10Aはクロックタイム2に対応するデータのすぐ次の4つのビットを次のスイッチ10Fに渡し始める。しかしながら、選択コマンドに続く情報ビット(本例ではクロックタイム2の4本のデータラインによって伝送される情報ビット)は、図8に示されるように常に全てゼロ(DEAD FIELD)でなければならない。この目的は後で説明する。

【0045】クロックタイム2において、スイッチ10Aの入力ポート1から出力ポート2への接続が行われ、クロックタイム2の信号シーケンスに、スイッチ10A及び相互接続配線を介してスイッチ10Fの入力ポート1に伝送させる。この時点から、スイッチ10Aは後続の全てのデータをスイッチ10Fの入力ポート1に直ちに伝送するだけで、その入力ポート1のインタフェースを介してスイッチ10Aに提供されるいかなる他のデータパターンの検査や処置も行わない。スイッチ10Aは、入力ポート1を介して受信する全てのデータパターンを出力ポート2及びスイッチ10Fに直ちに渡すだけである。従って、クロックタイム2において、スイッチ10A及びその関連するケーブルを介する遅延がゼロであると仮定すると、スイッチ10Fの入力ポート1は、有効(VALID) 信号が立ち上がり、4本のデータラインの全てのゼロデッドフィールド(DEAD FIELD)がスイッチ10Fの入力ポート1に入るのを見る。このようにして、10Aの入力ポート1がタイム0で以前に使用可能にされたのと同じように、タイム2においてスイッチ10Fの入力ポート1が使用可能にされる。

【0046】本例では、INI-DATA3はクロックタイム3の間に論理1に進むことによって、スイッチ10Aがクロックタイム1の間に入力ポート1から出力ポート2に接続するよう命じられたのと同じように、スイッチ10Fは入力ポート1から出力ポート3に接続するよう命じられる。スイッチ10Fは命令された接続を行うときに、クロックタイム3においてデータラインから4個のビットを除去し、図5の制御ブロック50Aの一部である選択レジスタに記憶する。クロックタイム3の間に伝送されるビットはスイッチ10Fを介してノード7に渡されず、代わりにスイッチ10Fがクロックタイム4に

対応するデータのすぐ次の4つのビットをノード7に渡し始める。しかしながら、選択コマンドに続く情報ビット(本例ではクロックタイム4の4本のデータラインによって伝送される情報ビット)は、図8に示されるように常に全てゼロ(DEAD FIELD)でなければならない。従って、クロックタイム4までに、スイッチ10A及び10Fはノード1からノード7へ直接データを転送するための接続経路を設定した。クロックタイム5まで、ノード7は遊休(IDLE)コマンド以外は何も見えていない。タイム4において、ノード7はスイッチ10FからのOUT3-VAL IDラインが作動することによって、タイム5でデータを受信し始めるのが可能になる。タイム5から、ノード7はスイッチ10Fからの4本のOUT3-DATAラインを介してノード1からデータを受信することができる。伝送される実際のデータのプロトコルは、プリアンプル等でコード化しているマンチェスターコード化された8/10ビット等のいかなる通常のフォーマットであってもよい。しかしながら、好ましい実施例において、図8に示されるように、NRZデータメッセージが続く、タイム5における同期フィールド全て1である。データメッセージは、転送のワードカウント長を特定できる。全て1の同期フィールドを実際のデータメッセージの接頭語とする目的は、1クロックタイムにおいて受信ノード7を送信ノード1に同期させることである。これは、データ転送に含まれる2つのノードが互いに非同期的なクロックシステムを有するが、特定の許容範囲内の同じ周波数で作動していると仮定する。

【0047】好ましい実施例は、クロックタイム6とクロックタイム7の間で最初にメッセージのワードカウント長を伝送することである。次に、ノード7は長さカウントを減らし、転送が完了する時を検出することができる。ノード7は次に選択されたエラー検出方法(パリティ、ECC又はCRC)を使用してメッセージの正確さを検査することができる。メッセージが正確に受信されるならば、ノード7はクロックタイム $n+1$ 及び $n+2$ で受容(ACCEPT)インタフェースラインをスイッチ10Fに作動させることによって応答する。スイッチ10Fはスイッチ10Aに受容(ACCEPT)表示を渡し、またノード1に直ちに受容(ACCEPT)表示を戻す。これは、転送が無事に完了したことをノード1に示して、ノード1がスイッチ10Aへの有効(VALID)ライン及び4本のデータラインをゼロにリセットして、データ転送を完了して遊休(IDLE)状態に戻る。スイッチ10AへのINI-VALID入力ラインはタイム $n+3$ でゼロになり、スイッチ10Aの入力ポート1に出力ポート2への接続を破壊させ、遊休(IDLE)状態に戻る。直ちに、スイッチ10FはINI-VALID入力ラインがゼロになるのを見て、出力ポート3への接続を破壊し、遊休(IDLE)状態に戻る。従って、接続は破壊されて、スイッチは1クロックタイム程の少なさで遊休(IDLE)状態に戻ることができる。ノード1が伝送すべきも

う1つのメッセージを有するならば、次のメッセージをバッファ52及びシフトレジスタ54(図6)にロードし、タイム $n+4$ になるとすぐにノード7又は他のいかなるノードへと伝送し始める。ノード1によって生成される有効(VALID)信号は、もう1つの転送が始まる前の1つの転送の終了を示すために、最低1クロックタイム(タイム $n+3$)でゼロに戻らなければならないことが唯一の制約である。

【0048】ワードカウントがクロックタイム n でゼロになった後で、ノード7が受信したメッセージにエラーがあることを発見したならば、(受容(ACCEPT)の代わりに)拒否(REJECT)インタフェースラインをスイッチ10Fに作動することによって応答する。スイッチ10Fはノード7から入力する拒否(REJECT)信号を使用して、ノード7への接続を破壊し、拒否(REJECT)表示をスイッチ10Aに渡し、また、接続を破壊して遊休(IDLE)状態に戻った後でノード1に拒否(REJECT)表示を直ちに戻す。次にノード1は転送が拒否されたことを知らせ、スイッチ10Aへの有効(VALID)ライン及び4本のデータラインをゼロにリセットすることによってIDLE状態に戻る。ノード1は次にバッファ52からシフトレジスタ54を再ロードし、最初(クロックタイム-1)から再び伝送を開始することによって伝送を再度試みる。再伝送は以前に拒否された伝送と同一の経路を介して生じるか、或いは、ネットワークを介して他の経路が実行されるならば、もう1つの経路が試みられる。連続拒否(REJECT)が見られて、特定の数の拒否(REJECT)が同じメッセージで発生するならば、エラー報告メカニズムが呼び出されてもよい。

【0049】ネットワーク経路においていかなるスイッチ10もメッセージを拒否(REJECT)することができる。これは2つの場合のいずれにも発生する。

【0050】1) BUSY- もしスイッチが接続するように命じられる出力ポートがBUSY(即ち、以前に設定された接続によって使用されている)ならば、スイッチは拒否(REJECT)ラインを以前のネットワークステージ又は送信機に作動する(ネットワークの第1ステージがBUSYを検出する場合)ことによってコマンドを出す入力ポートにこのBUSY状態を示す。例えば、図8に示される例において、10Aがクロックタイム-2で入力ポート4から出力ポート2に接続するようなコマンドを受信したならば、入力ポート1がクロックタイム1で出力ポート2に接続するように要求されるとき作動する。この場合、出力ポート2はクロックタイム1においてBUSYであり、スイッチ10Aはノード1にINI-REJECTラインを作動する。上記のように、送信機はいかなる拒否(REJECT)されたメッセージも再度試みる。

【0051】同様に、接続はスイッチ10Aで無事に行われるが、スイッチ10Fの出力ポート3はクロックタイム3でBUSYとなり、スイッチ10Fにスイッチ10A

への拒否(REJECT)信号を出させる。また、これは10Aに接続を破壊し、遊休(IDLE)状態に戻った後で、ノード1へ直ちに拒否(REJECT)信号を戻す。

【0052】2) 同時競合(Simultaneous CONTENTION)-

上記のようにクロックタイム2でスイッチ10Aにおいて入力ポート4が出力ポート2と接続するよりも(クロックタイム1における入力ポート1からの同じコマンドに先立って)、2つ以上の入力ポートにはほぼ同時に同じ出力ポートを接続するように試みることは可能である。

これは、使用可能な出力ポートへの競合(CONTENTION)と称される。例えば、入力ポート1及び4の双方がクロックタイム1において、出力ポート2へ接続を要求する同時コマンドを送信したと仮定する。本発明は、競合する2つの入力ポート1及び4を出力ポート2へ最初に接続することによってこの競合を解決する。正味の効果は2つの入力ポートを出力ポート2に電氣的に接続することであり、2つのソースから来る信号を論理的にOR(オア)する。クロックタイム2の間、2つの入力ポート1及び4が同じために2つの入力ポートの論理的ORはエラーを生じない(NOT)。それぞれの有効(VALID)ラインは論理1であり、それぞれのデータラインはデッドフィールド(DEAD FIELD)(論理0)を含む。しかしながら、クロックタイム3において各ソースからの信号は異なってもよく、2つの入力ポートがクロックタイム3以後接続されたままであるならばエラーが生じる。換言すれば、スイッチ10Aは2つ以上の入力を同じ出力に接続するとして決定を補正するための1サイクルタイム

(クロックタイム2)を有する。スイッチ10Aは、1より多い入力が入力の出力に接続されるという事実を検出することによって、クロックタイム2の間でこの補正を行う。次に、1つの接続を除く全ての複数の接続をリセットすることによって、クロックタイム3になる前に処置を行う。どの接続をリセットしどの接続を保持するかは、優先順位に基づいて決定される。好ましい実施例において、簡単な優先順位の概要が以下のように用いられる。もし入力ポート1が競合しているならば、接続を得る。入力ポート1が競合せず入力ポート2が競合するならば、入力ポート2が接続を得る。もし入力ポート1及び2が競合せず、入力ポート3が競合するならば、入力ポート3が接続を得る。他の全ての入力ポートが接続を望まないときに限り、入力ポート4が接続を得る。この優先順位による選択を本例に適用すると、入力ポート1が出力ポート2への接続を保持すると共に、入力ポート4から出力ポート2への接続がクロックタイム2の間にリセットされる。これによって、スイッチ10Aから入力ポート4に通常の方法で出される拒否(REJECT)信号が生じる。

【0053】従って、本発明のデッドフィールド(DEAD FIELD)の目的は、スイッチステージあたりの1クロックタイムで同時競合を解決することである。デッドフィー

ルド(DEAD FIELD)の第2の目的は、前のクロックタイム間で作動していた選択ビットに立ち下がりエッジを生じ、直列選択データを連結されたスイッチに搬送する4本のデータラインを介して現れるタイミングスキュー(ずれ)を補償することである。スイッチに接続を行うように命じるデータの立ち上がり及び立ち下りの双方は、トリガして決定することができるクロック解除されたスイッチに2つのクロックエッジ(立ち上がり及び立ち下がり)を与える。これら2つのクロックエッジだけが、ALLNODE スイッチにある2つの決定時間である。

【0054】図9を参照すると、スイッチ方式マイクロチャネル平面(プレーナ)の物理的レイアウトがより詳細に示されている。スイッチ方式マイクロチャネル平面は、スイッチネットワークの1つのノード又はポートを形成する2つのカードのグループに編成された16個のマイクロチャネルカードスロット200、各々がノードあたり2個のマイクロチャネルカードを固有のMCC機能へ接続する8つの固有のマイクロチャネルバス500、8つのマイクロチャネルバスをAllnode スイッチインタフェースに適用する8個のマイクロチャネル変換装置(MCC)130、ネットワーク30機能の全体又はネットワーク30機能の一部を形成する2つの8x8Allnode スイッチチップ360及び350、並びに、平面コネクタ700及び800をそれぞれ介してブロック360及び350からのスイッチインタフェースを駆動及び受信し、単一のスイッチ方式マイクロチャネル平面100の一部として実行されない残りのスイッチネットワーク30へ通信及び拡張するために使用されるシングルエンドドライバ回路400及びシングルエンドレシーバ回路450を実行する。ブロック360及び350で示されるようなAllnodeスイッチの8x8バージョンは、本明細書に記載される基本的な4x4Allnodeスイッチの単純な拡張であることに注意したい。4x4スイッチが4個の入力ポート及び4個の出力ポートを有するのと比べ、8x8スイッチは8個の入力ポート及び8個の出力ポートを有するのが唯一の違いである。8x8スイッチは、いかなる入力ポートと8個全ての出力ポートの間に完全な相互接続能力を提供する。

【0055】図10は、2ステージスイッチネットワーク30機能全体を形成し、単一のスイッチ方式平面100が完全な8ノードシステムとして独立することを可能にするのに必要な外部平面接続701乃至708を示す。この場合、スイッチブロック360及び350はネットワーク30の機能全体を形成し、スイッチ360が2ステージネットワークの第1ステージであり、スイッチ350が第2ステージである。図10は、スイッチ360の8個の出力が、ジャンパワイヤ701乃至708及びドライバ400及びレシーバ450を介して、スイッチ350の8個の入力へ外部でジャンパされる方法を示している。ジャンパ701乃至708は8つの多重経

路を形成し、8つのジャンパの内のいかなる1つも、いかなる8個のノードによっていかなる他の8個のノードに達する経路として使用できる。

【0056】例えば、ノード1からノード8への、1つのノードからもう1つのノードへの一般的なデータ転送は以下のステップを含む。送信ノード1は、標準的マイクロチャンネルバスシーケンスを用いてそのローカルマイクロチャンネルバス500と通信を行う。MCC130はこれらのシーケンスに応答し、シーケンスをインタフェース600を介して通信されるスイッチネットワーク30によって理解されるシーケンス及びプロトコルに変換する。従って、第1ノードの2つのMCカードの内の1つからの通信はMCバス500を介してMCC130に渡され、MCC130が通信フォーマットを変換して、スイッチネットワーク30に渡す。スイッチネットワーク30はMCC130によって命令されて受信ノード8に接続を行い、ネットワーク30が通信を渡して、スイッチ360及びドライバ400を介してスイッチネットワーク出力バス700へ送られる。通信は、8つのジャンパ701乃至708のいかなる1つによってネットワークの平面でない(off-planar)部分を介して進み、スイッチネットワーク入力バス800を介して平面100に戻る。通信データは、シングルエンドレシーバ450を通過し、スイッチ350及びスイッチインタフェース607を介して特定の受信ノードに送られる。次に、データはMCC137によって受信され、標準的MCフォーマットに変換し直されて、マイクロチャンネルバス507を介して受信ノード8に渡される。

【0057】スイッチ方式マイクロチャンネル平面はモジュール的であり、2つの方法のいずれかによって8ノードより多い拡張性を可能にする。スイッチ方式マイクロチャンネル平面の概念は、複数のスイッチ方式マイクロチャンネル平面を使用することによってより大型のシステムに拡張できる。この例は図11及び図12に示され、ここで8個のスイッチ方式マイクロチャンネル平面の相互接続が、2つのスイッチステージを介して相互接続される64ノードシステムを実行するように示されている。図11及び図12では、スイッチブロック360及び350がドライバ400及びレシーバ450(簡素化のため図11及び図12に図示せず)を介して外部接続を有する唯一のブロックであるために、8個の平面の各々からの2つのスイッチブロック360及び350だけが示されている。これらの外部接続は、2ステージネットワークを形成するために互いに配線される。従って、8個までの同一のスイッチ方式マイクロチャンネル平面100が相互接続されて、8個の平面に完全に含まれるスイッチングネットワークを介して相互接続される128個のマイクロチャンネルカード、並びに、16個のチップから成る、64ノードシステムを提供する。

【0058】更に、図13に示されるように、追加のス

イッチングがスイッチ方式マイクロチャンネル平面の完全な外部で実行されて、複数のスイッチ方式マイクロチャンネル平面を補足するために使用される。中心にあり、いかなるスイッチ方式マイクロチャンネル平面にも含まれないスイッチネットワーク30の補足的な部分を導入することによって、8個より多いスイッチ方式マイクロチャンネル平面を互いに相互接続することができる。この概念を用いて、システムを構成するノード及びマイクロチャンネルカードの数が理論上の限度なしに拡張される。

【0059】マイクロチャンネル変換装置(MCC)には、マスター又はスレーブの2つの可能な実施がある。これらマイクロチャンネルの機能的詳細は、他の同時出願に詳細に記載され、参照のため本明細書に組み込まれる。

【0060】米国特許出願第07/946、509号、「マイクロチャンネルバスからスイッチに変換するためのマスター変換装置(MASTER CONVERTER APPARATUS FOR CONVERTING MICROCHANNEL BUS TO SWITCH)」

米国特許出願第07/946、986号、「マイクロチャンネルバスからスイッチに変換するためのスレーブ変換装置(SLAVE CONVERTER APPARATUS FOR CONVERTING MICROCHANNEL BUS TO SWITCH)」

【0061】

【発明の効果】本発明は上記より構成され、複数のバス方式プロセッサカードと複数の拡張カードの少なくとも一方に挿入し、本発明の平面上にあるマルチステージスイッチネットワークを介して前記カードを相互接続するためのモジュール的に拡張可能なスイッチ方式平面装置が提供される。

【図面の簡単な説明】

【図1】複数のバス方式プロセッサと複数の拡張カードの少なくとも一方を有するスイッチ方式マイクロチャンネル平面装置の、本発明の好ましい実施例を概して示している。

【図2】4個までのノードを相互接続するために完全に並列な交換手段を提供する能力を有する4入力及び4出力(4x4)クロスバー交換装置を示す。

【図3】4個のノードの中で完全な相互接続を提供するために、4入力及び4出力(4x4)クロスバー交換装置に必要な相互接続を示す。

【図4】4個より多いノードを有するシステムを許容するための本発明の交換装置に開示された4x4の実施例を連結するための一般的な方法を示す。

【図5】本発明で使用して、デジタルデータを転送するための4個までのシステムノードを相互接続する完全な並列交換手段を提供する能力を有する、4x4クロスバー交換装置の単純なデジタルデータフローと制御経路の実行の概略ブロック図を示す。先行の米国特許出願第07/677、543号に予め開示されている。

【図6】4本の同期データラインを介して本発明の交換

装置の先の4×4の親実施例に送信される並列制御と多重ライン直列デジタルデータ情報を生成するための一般的な方法を示す。

【図7】本発明の交換装置の先の4×4の親実施例の1つの入力ポートに到着するデジタルインタフェース信号を1つの出力ポートに回すための一般的なタイミング図である。

【図8】一方のノードからもう一方のノードにデジタルデータを送信するために、本発明の交換装置から成るネットワークを介する伝送経路を選択及び設定する一般的な方法を示す。

【図9】スイッチネットワークが平面に組み込まれる方法を示しているスイッチ方式マイクロチャネル平面装置の一般的な物理的実施例を示す。

【図10】スイッチ方式マイクロチャネル平面装置の8個のノードを独立したシステムとして相互接続するのに必要な平面相互接続を示す。

【図11】各々がより多数のノードを相互接続できる分配されたマルチステージネットワークの一部を組み込んだ、複数のスイッチ方式マイクロチャネル平面を相互接続することによって、システムにおけるノードの数が8ノードより多く拡張される方法を示す。

【図12】各々がより多数のノードを相互接続できる分

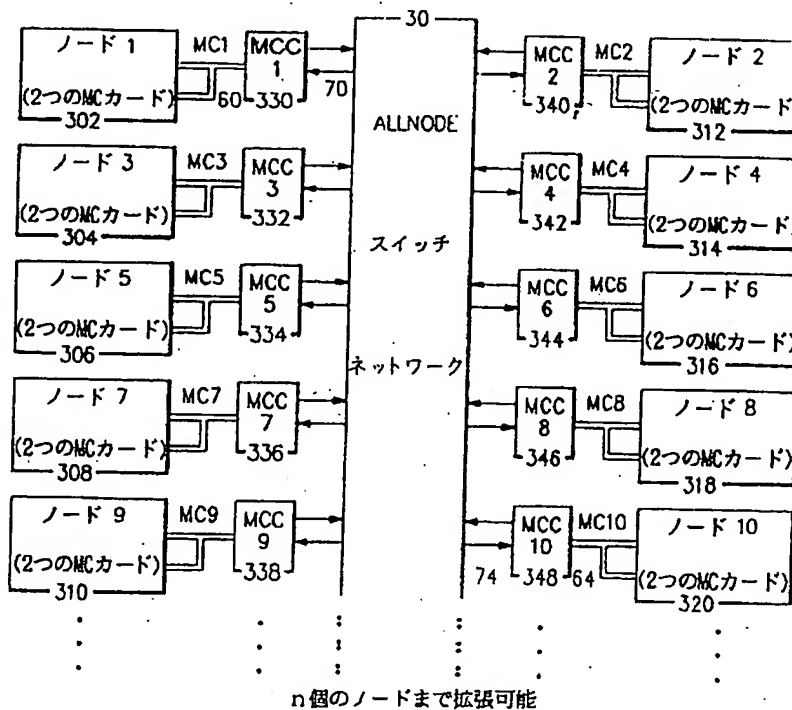
配されたマルチステージネットワークの一部を組み込んだ、複数のスイッチ方式マイクロチャネル平面を相互接続することによって、システムにおけるノードの数が8ノードより多く拡張される方法を示す。

【図13】複数のスイッチ方式マイクロチャネル平面を補足するスイッチネットワークへ相互接続することによって、システムにおけるノードの数が64ノードより多く拡張される方法を示す。ここで、スイッチネットワーク全体は、複数のスイッチ方式マイクロチャネル平面及び補足するスイッチネットワークを介して分配される。

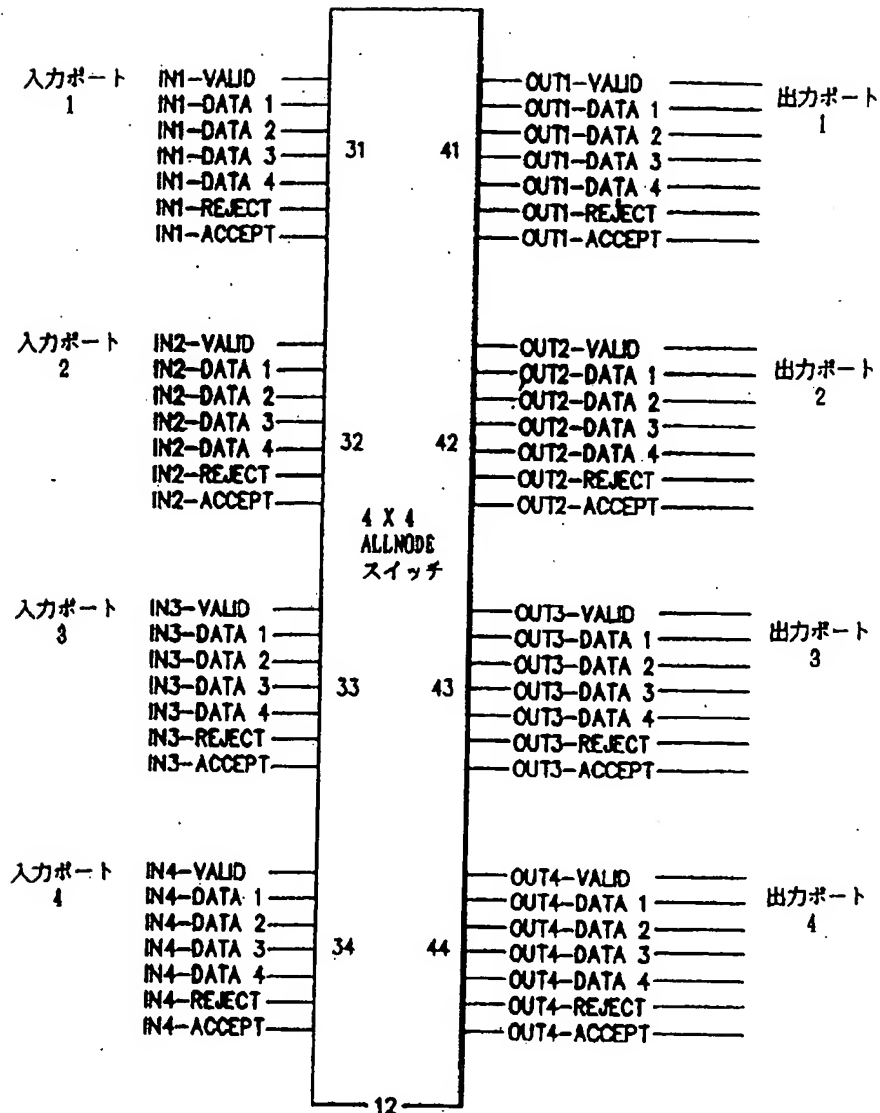
【符号の説明】

- 30 スイッチネットワーク
- 100 スイッチ方式マイクロチャネル平面
- 130、137 MCC
- 200 マイクロチャネルカードスロット
- 350、360 スイッチ
- 400 シングルエンドドライバ
- 450 シングルエンドレシーバ
- 500、507 MCバス
- 600、607 インタフェース
- 700、800 平面コネクタ
- 701乃至708 ジャンパワイヤ

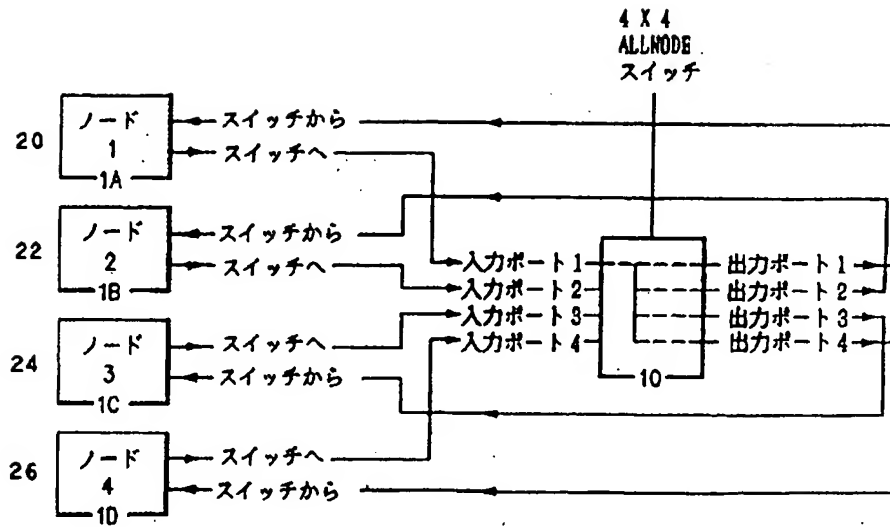
【図1】



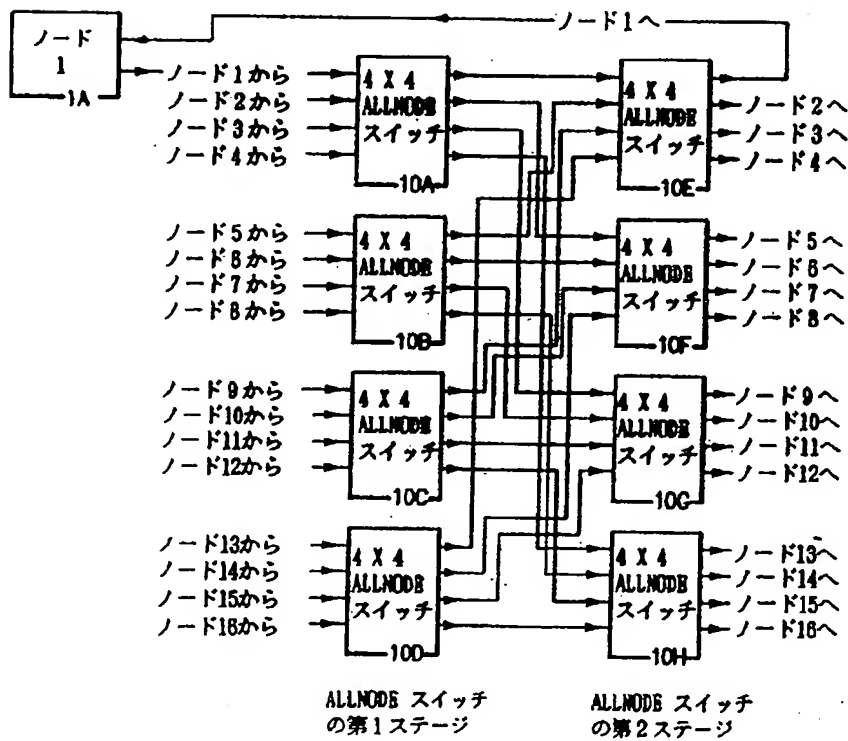
【図2】



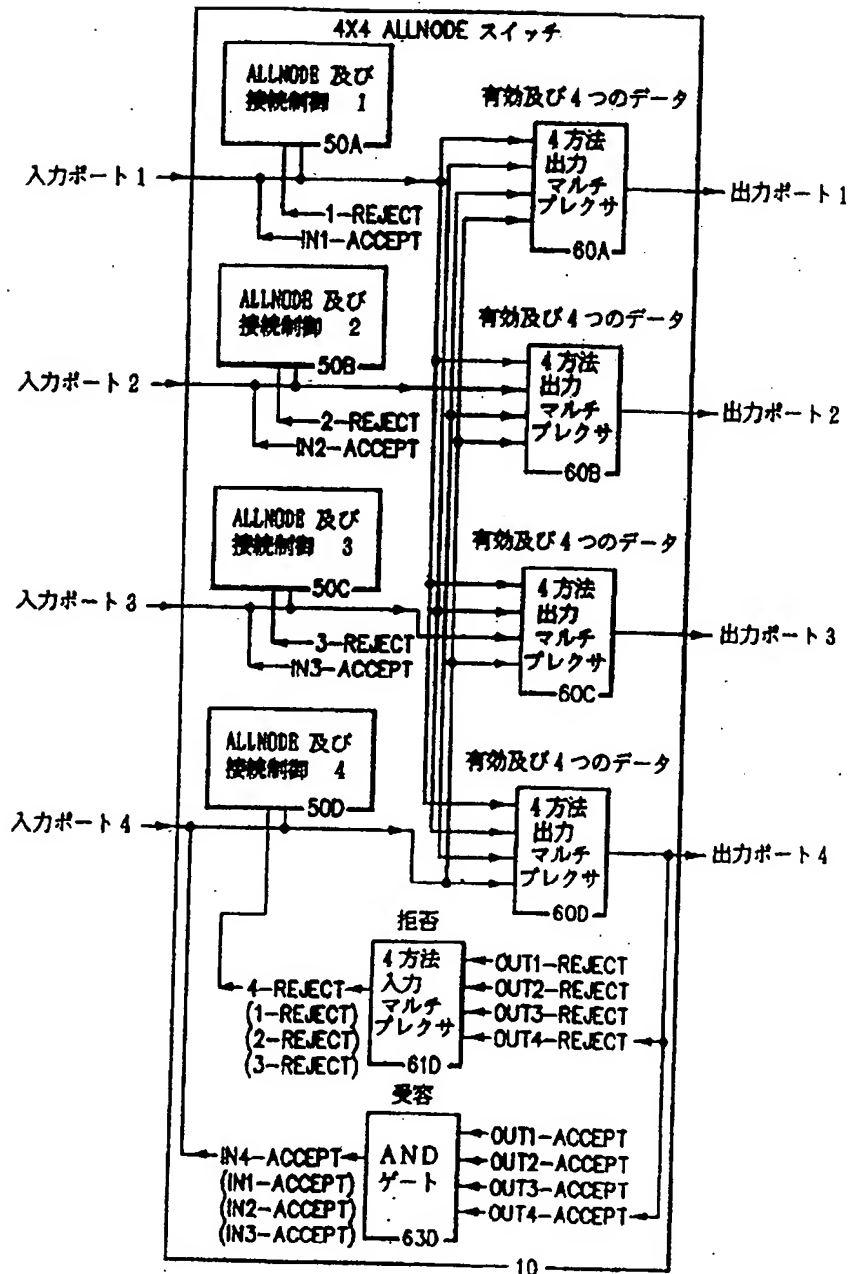
【図3】



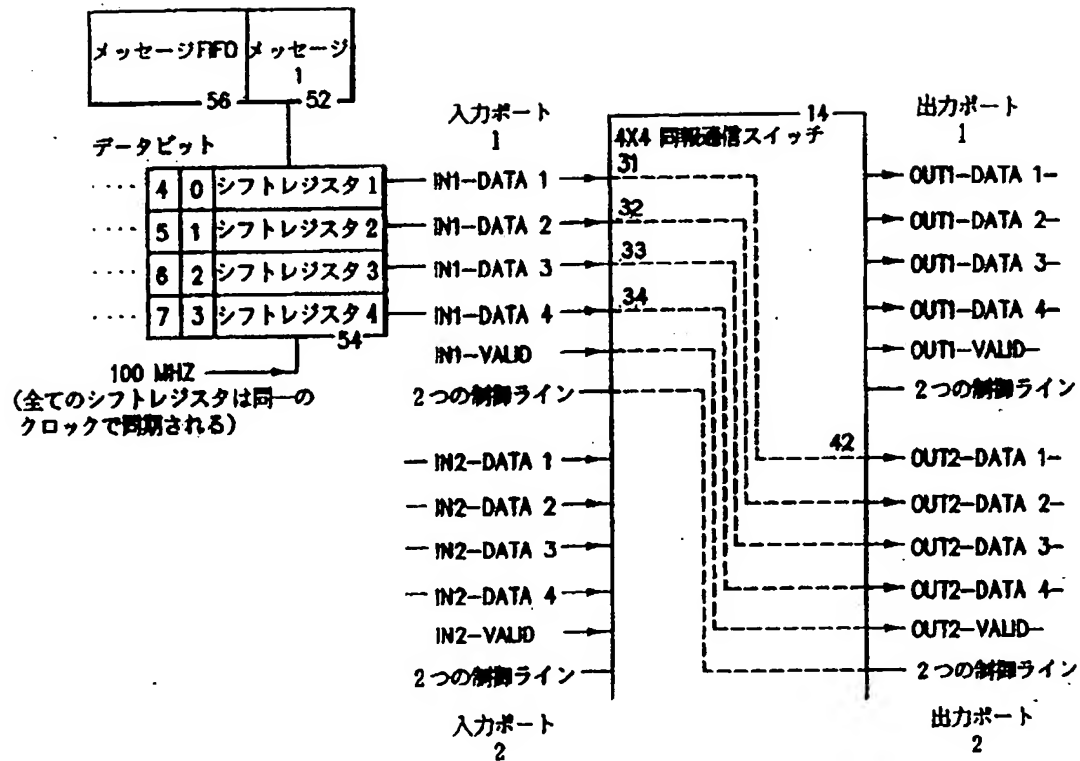
【図4】



【図5】

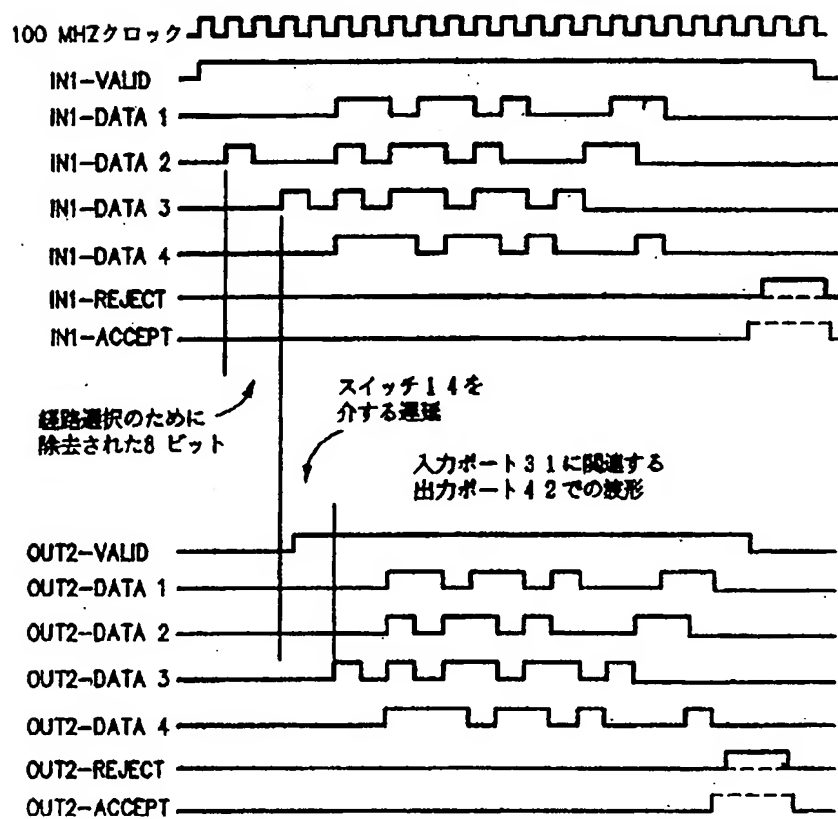


【図6】

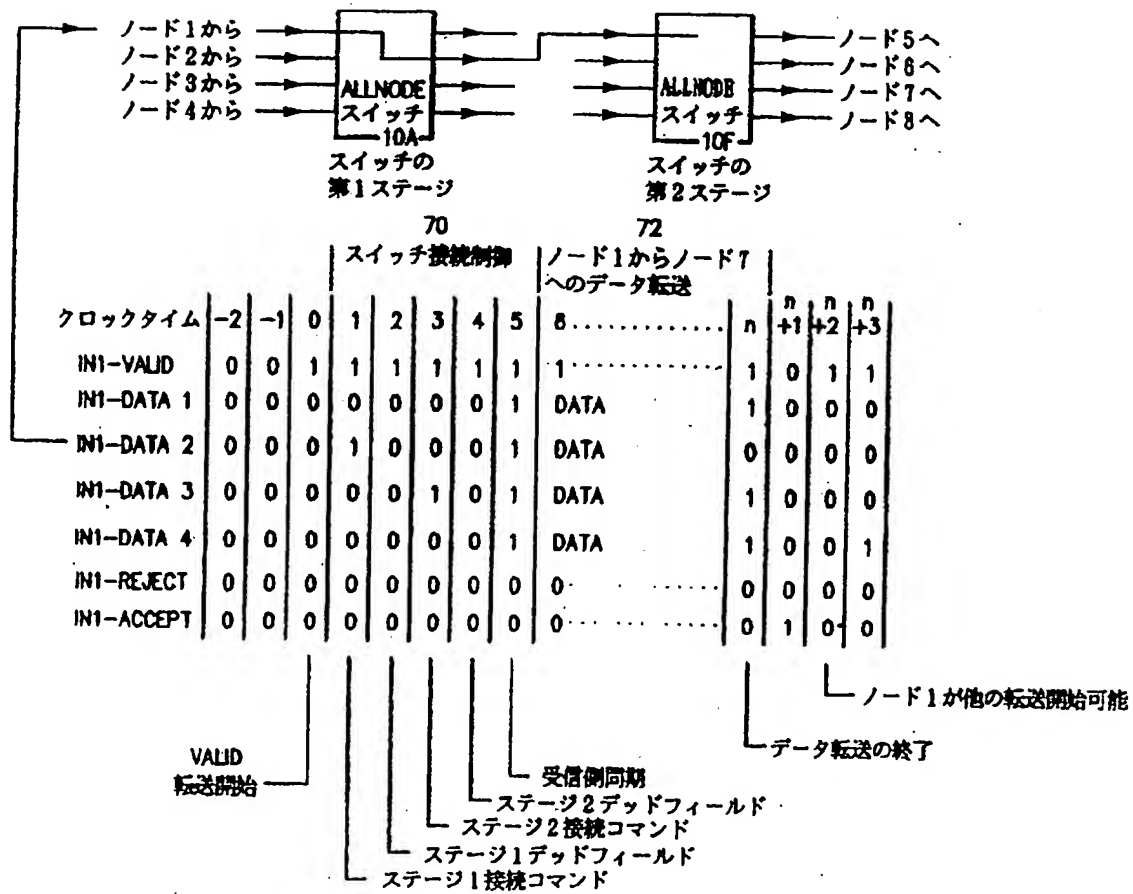


【図7】

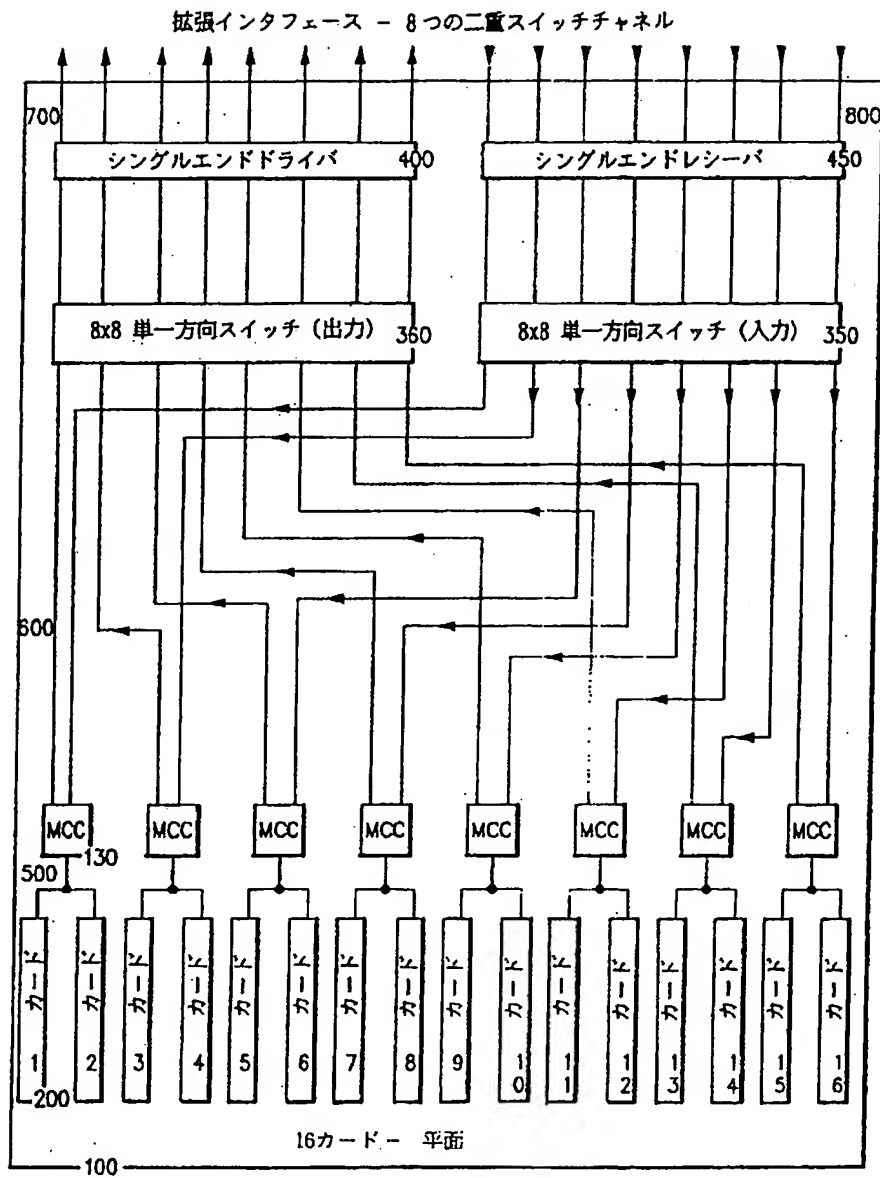
シフトレジスタ54の100MHzクロックに関連する
スイッチ入力ポートインタフェース31での波形



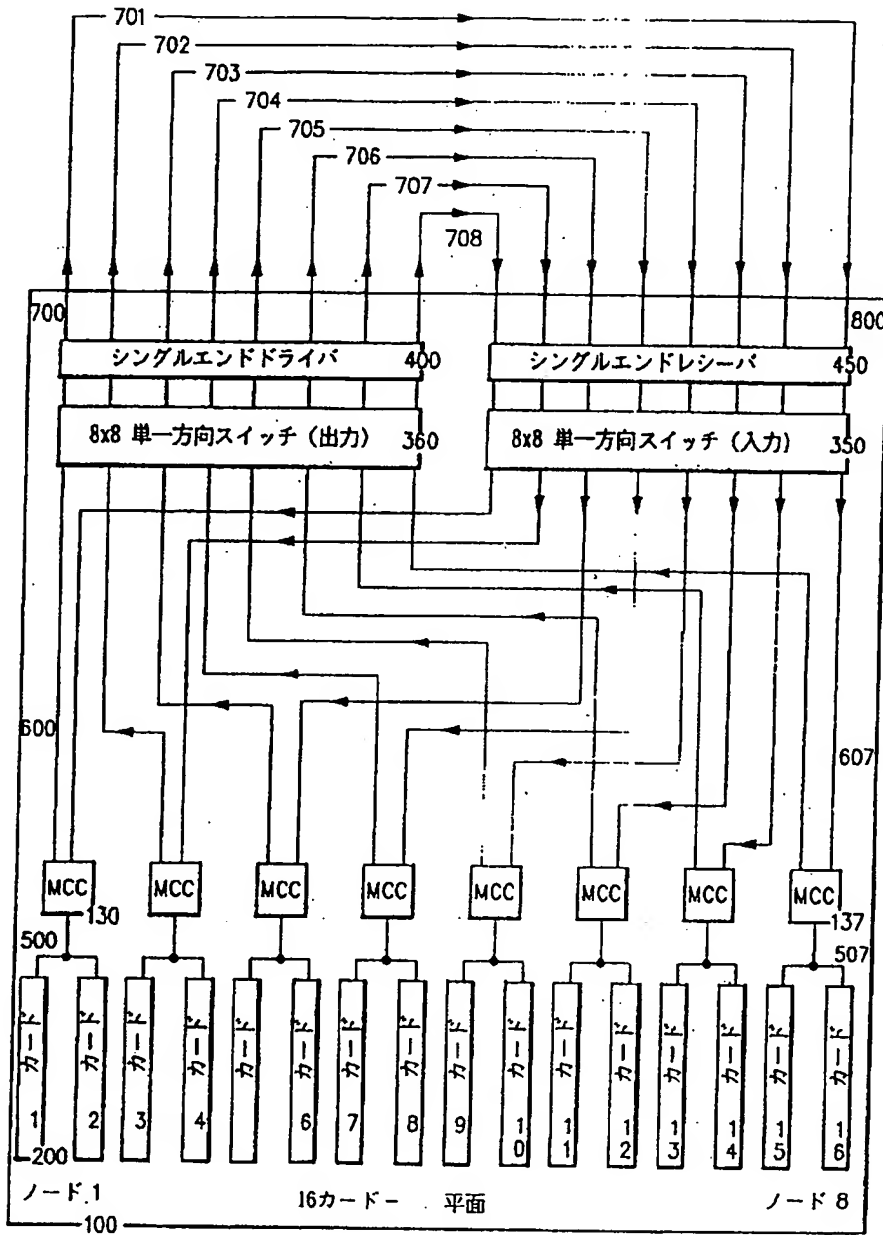
【図8】



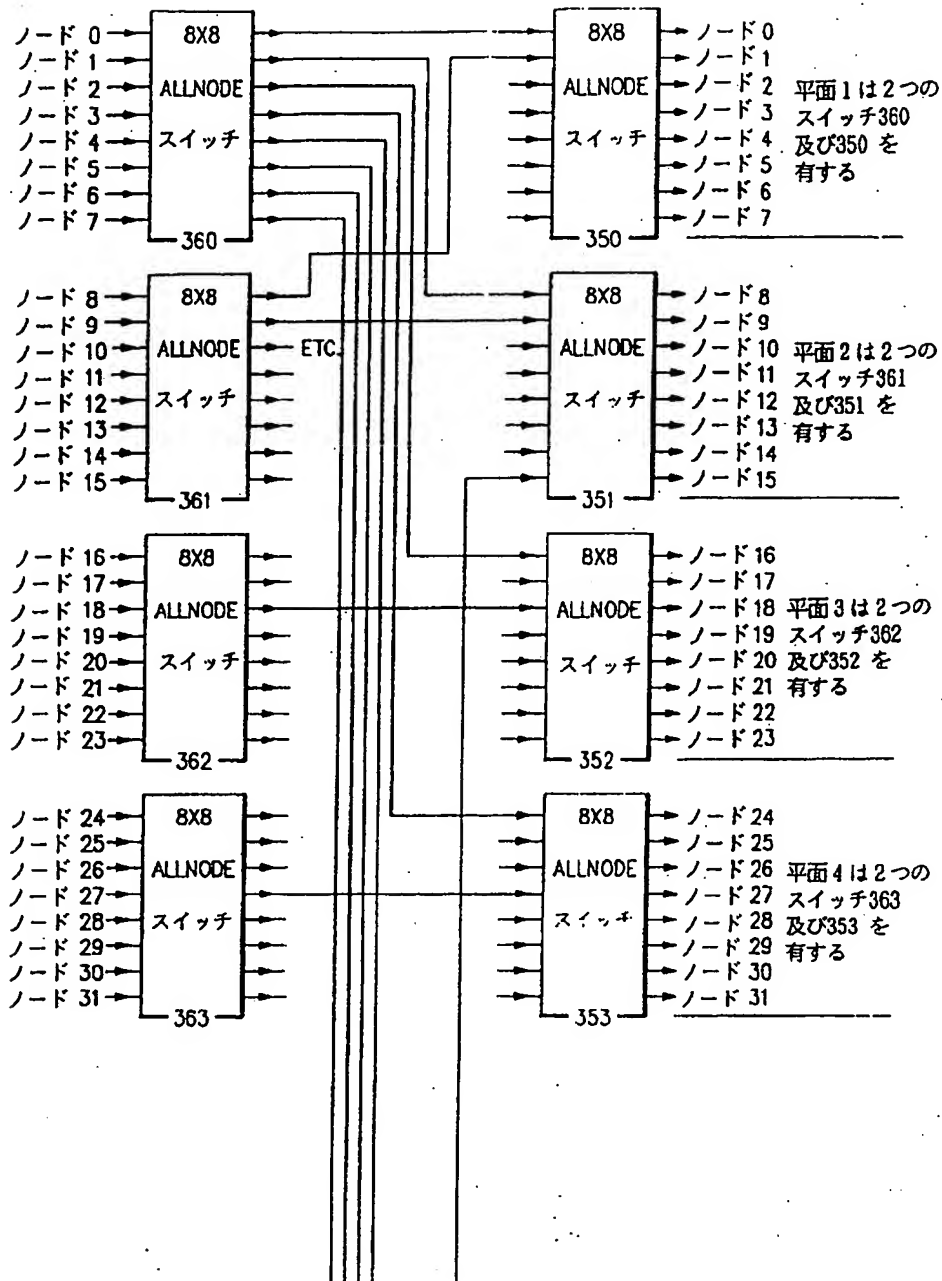
【図9】



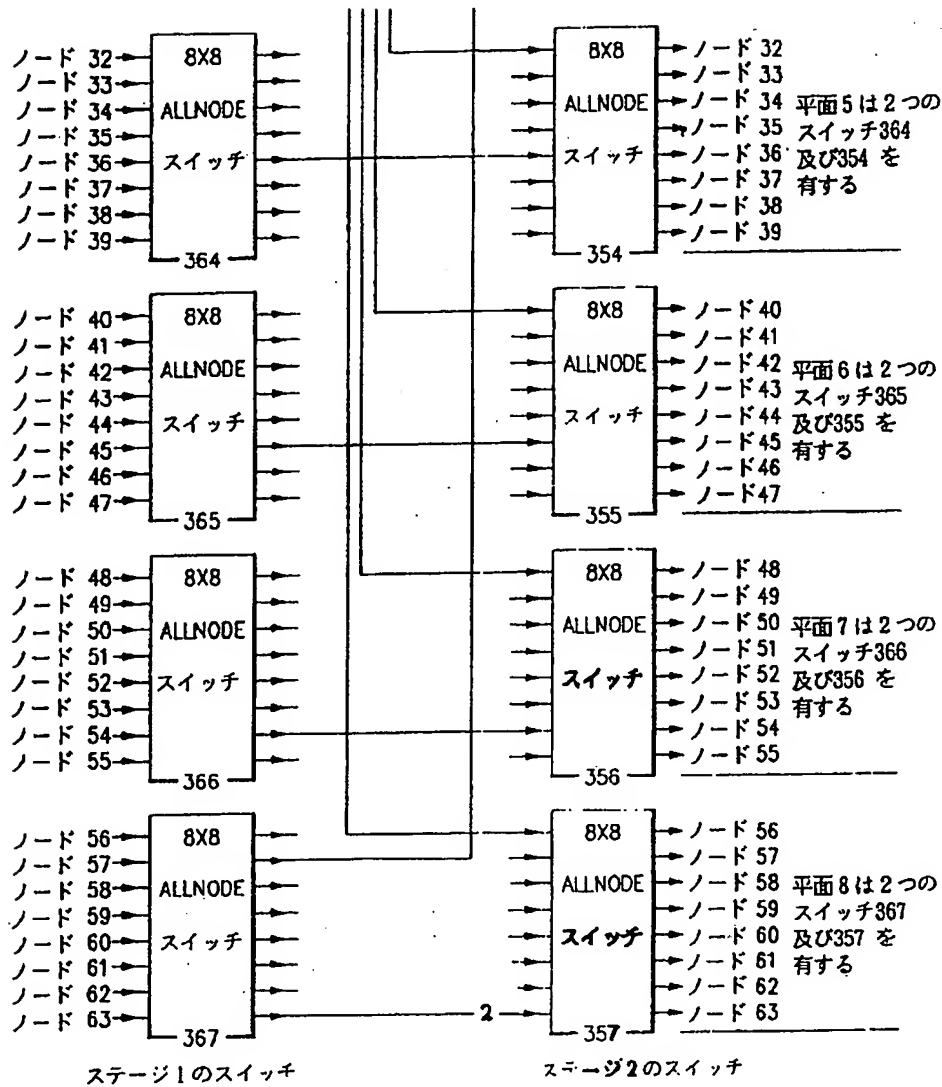
【図10】



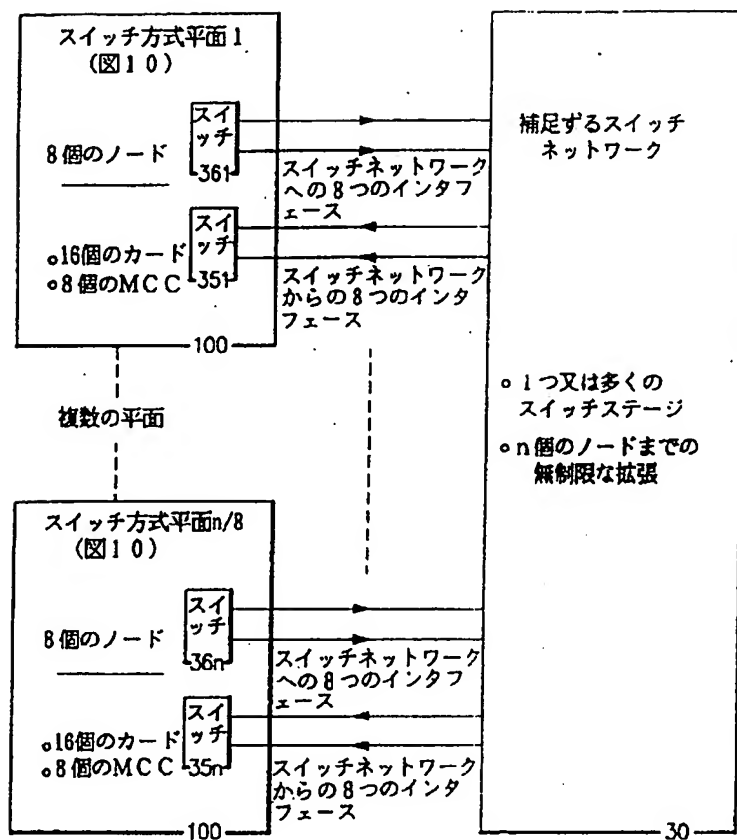
【図11】



【図12】



【図13】



フロントページの続き

(72)発明者 マイケル ハンス フィッシャー
アメリカ合衆国55901、ミネソタ州ロチェ
スター、コッパーフィールド レイン
4442

(72)発明者 ロバート フランシス ラッシュ
アメリカ合衆国13850、ニューヨーク州ヴ
ェスタル、コートランド ドライヴ 3100

(72)発明者 マイケル アンソニー マニグエット
アメリカ合衆国13827、ニューヨーク州オ
ウゴ、ホリスター ロード 223

(72)発明者 オマー アジズ サイード
アメリカ合衆国22094、バージニア州レス
トン、クウェイル リッジ ロード
11618